electronic reprint

Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449
Editors: **E. N. Baker and Z. Dauter**

Mycobacterium tuberculosis RmlC epimerase (Rv3465): a promising drug-target structure in the rhamnose pathway

Katherine A. Kantardjieff, Chang-Yub Kim, Cleo Naranjo, Geoffry S. Waldo, Timothy Lekin, Brent W. Segelke, Adam Zemla, Min S. Park, Thomas C. Terwilliger and Bernhard Rupp

Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Mycobacterium tuberculosis RmlC epimerase (Rv3465): a promising drug-target structure in the rhamnose pathway

Katherine A. Kantardjieff,[a] Chang-Yub Kim,[b] Cleo Naranjo,[b] Geoffry S. Waldo,[b] Timothy Lekin,[c] Brent W. Segelke,[c] Adam Zemla,[c] Min S. Park,[b] Thomas C. Terwilliger[b] and Bernhard Rupp[c]*

[a]Department of Chemistry and Biochemistry and W. M. Keck Foundation Center for Molecular Structure, California State University Fullerton, Fullerton, CA 92834, USA, [b]Bioscience Division, MS M888, Los Alamos National Laboratory, Los Alamos, NM 87545, USA, and [c]Biology and Biotechnology Research Program, L-448, University of California, Lawrence Livermore National Laboratory, Livermore, CA 94551, USA

Correspondence e-mail: br@llnl.gov

Received 17 February 2004
Accepted 5 March 2004

PDB Reference: dTDP-4-keto-6-deoxyglucose epimerase, 1upi, r1upisf.

The *Mycobacterium tuberculosis rmlC* gene encodes dTDP-4-keto-6-deoxyglucose epimerase, the third enzyme in the *M. tuberculosis* dTDP-L-rhamnose pathway which is essential for mycobacterial cell-wall synthesis. Because it is structurally unique, highly substrate-specific and does not require a cofactor, RmlC is considered to be the most promising drug target in the pathway, and the *M. tuberculosis rmlC* gene was selected in the initial round of TB Structural Genomics Consortium targets for structure determination. The 1.7 Å native structure determined by the consortium facilities is reported and implications for *in silico* screening of ligands for structure-guided drug design are discussed.

1. Introduction

The TB Structural Genomics Consortium (TBSGC) is one of the nine NIGMS-funded Protein Structure Initiative Pilot projects serving as a structural biology resource for the *Mycobacterium tuberculosis* (MTB) research community (Terwilliger *et al.*, 2003). Consortium members can target proteins of interest, and highly ranked targets are produced at the Los Alamos protein-production facility and shipped to the crystallization facility at Lawrence Livermore National Laboratory (LLNL) for automated high-throughput crystallization (Rupp *et al.*, 2002). Data are collected at the Advanced Light Source (ALS) in Berkeley, and the structures are determined by the LLNL or ALS beamline members. Coordinates are deposited within three weeks of the final refinement.

TB is a re-emerging disease with an increasing prevalence of multi-drug resistant strains (Ramaswamy & Musser, 1998), and a long-term goal of the TBSGC is to provide a foundation for structure-guided drug design. Protein targets of high priority are those which are essential or unique to the bacillus. Mycobacterial cell-wall biosynthesis, the target of the well known drugs isoniazid and ethambutol (Schroeder *et al.*, 2002), has been of particular interest in the development of antimycobacterial therapeutics (Kantardjieff & Rupp, 2004). Rhamnose-synthetic enzymes are attractive targets in cell-wall synthesis. L-Rhamnose, a sugar that is not present in the human host, plays a key role as a structural link between the mycobacterial cell-wall components arabinogalactan and peptidoglycan. L-Rhamnose is derived from a precursor, dTDP-L-rhamnose, which is synthesized from glucose-1-phosphate and dTTP by means of the dTDP-L-rhamnose pathway. The prokaryotic dTDP-L-rhamnose pathway consists of four enzymes, the three-dimensional structures of each of which have been determined from different bacteria.

RmlC, the third enzyme in the dTDP-L-rhamnose pathway, functions as an epimerase, converting dTDP-4-keto-6-deoxy-

Table 1
Data-collection and refinement statistics.

Values in parentheses are for the highest resolution bin (1.75–1.70 Å).

Data collection	
Space group	$P3_221$
Wavelength (Å)	1.000
Unit-cell parameters	
a, b (Å)	64.91
c (Å)	87.20
Resolution (Å)	55.90–1.7
Unique reflections	23837 (1839)
Redundancy	4.4 (4.0)
Completeness	99.2 (93.8)
R_{sym}	0.052 (0.410)
$\langle I/\sigma(I)\rangle$	46.8 (2.5)
Reflections with $I/\sigma(I) > 3$ (%)	87.8 (53.3)
No. molecules in AU	1
V_M (Å^3 Da^{-1})	2.14
Solvent content (%)	42.7
Refinement	
R_{free} value, random 5% of reflections	0.249 (0.340)
R value	0.201 (0.272)
R.m.s.d. bond lengths† (Å)	0.021
R.m.s.d. bond angles† (°)	1.786
Overall coordinate error‡ (Å)	0.127
RSCC (*Shake&wARP*)§	0.92
RSCC (*REFMAC*5)¶	0.96
Ramachandran appearance††, residues in	
Most favored region	155 (89.6%)
Additional allowed	17 (9.8%)
Generously allowed	1 (0.6%)
Disallowed	0

† Deviations from restraint targets (Engh & Huber, 1991). ‡ Estimated standard uncertainty: diffraction precision index (DPI) based on R_{free} (Cruickshank, 1999). § Real-space correlation coefficient, F_c map against averaged and weighted *Shake&wARP* map. ¶ Real-space correlation coefficient, F_o map against F_c map, as reported by *REFMAC*5. †† Regions as defined in *PROCHECK* (Laskowski *et al.*, 1993).

glucose to dTDP-4-keto-rhamnose (Stern *et al.*, 1999). Because it is structurally unique, highly substrate-specific and does not require a cofactor, RmlC is considered to be the most promising drug target in the pathway, and the MTB *rmlC* gene was selected in the initial round of consortium targets for structure determination. The structures of RmlC from *Methanobacterium thermoautotrophicum*, *Salmonella typhimurium* and *Streptococcus suis*, both uncomplexed and bound to substrate analogs, have also been determined recently (Christendat *et al.*, 2000; Giraud *et al.*, 2000; Babaoglu *et al.*, 2003). We present a brief structure description, a comparison with the other uncomplexed apo structures and results from virtual ligand-screening studies of known and potential inhibitors. The coordinates (PDB code 1upi) have been deposited and released immediately in accordance with NIH guidelines for Structural Genomics Pilot Projects.

2. Experimental methods

2.1. Cloning and expression

A 0.6 kbp DNA fragment containing the *rmlC* gene (EMBL locus MTY13E12, accession No. Z95390.1, Rv3465) was amplified from *Mycobacterium tuberculosis* H37Rv genomic DNA as the PCR template, using the following oligonucleotide primers: 5′-AGATATA**CATATG**AAAGCACGCGAACTCGACGTCCCC-3′ and 5′-AATTC**GGATCC**GGTGCCGCGCATCTCCCCAATGAA-3′. The bases in bold represent *Nde*I and *Bam*HI sites, respectively. The amplified DNA fragment was digested with *Nde*I and *Bam*HI restriction enzymes and subcloned into the corresponding restriction sites in a modified pET28b vector which provided an N-terminal six-His tag upstream of the *Nde*I site. The expressed protein thus has the N-terminal extension MGSSHHHHHHSSGLVPRGSH and an additional GSV at the C-terminus.

Escherichia coli BL21 (DE3) cells were transformed with the *rmlC*-modified pET28b/His vector and grown to exponential phase at 310 K in 5 ml EZMix LB broth medium (Sigma) containing 30 µg ml^{-1} kanamycin and 50 µg ml^{-1} spectinomycin. This seed culture was transferred to 0.5 l EZMix Terrific broth medium (Sigma) and expression was induced with 0.5 m*M* IPTG at an OD_{600} of approximately 0.5. Growth was continued at 293 K for approximately 21 h until the OD_{600} reached approximately 15 (as inferred from dilutions). The cells were harvested and stored at 193 K.

2.2. RmlC purification

The cell pellet was lysed by sonication in 10 ml buffer *A* (20 m*M* Tris pH 8.0, 100 m*M* NaCl) per gram of cells for 10 min in 30 s pulses at 283 K. The cell debris was removed by ultracentrifugation for 30 min at 38 000 rev min^{-1} using a Ti-60 rotor (Beckman). The clear supernatant was filtered through a 0.2 µm pore membrane and loaded onto a 5 ml Talon superflow affinity column equilibrated with buffer *A*. After washing with 50 ml buffer *A*, the His-tagged RmlC was eluted from the cobalt-affinity column using buffer *B* (20 m*M* Tris–HCl pH 8.0, 500 m*M* NaCl and 300 m*M* imidazole). The elutant was dialyzed against buffer *C* (20 m*M* Tris–HCl pH 8.0, 100 m*M* NaCl, 10 m*M* β-mercaptoethanol) and purified by gel filtration on a Superdex-75 column using buffer *C* for equilibration and elution (Amersham Pharmacia Biotech). The peak fractions (monitored at OD_{280}) were analyzed by SDS–PAGE and the pooled protein fractions were concentrated to 26 mg ml^{-1} using a Centriprep YM-3 (Millipore). Protein, the purity of which was estimated to be 99% by SDS–PAGE and MALDI–TOF mass spectroscopy (Applied Biosystems), was stored at 277 K and shipped to the TB consortium crystallization facility at LLNL (Rupp *et al.*, 2002).

2.3. Crystallization

Crystals were grown in Greiner 96-well plates from sitting drops consisting of 2 µl protein stock solution mixed with 2 µl well solution. Conditions were screened using the CRYSTOOL random-screening protocol (Segelke, 2001) and the first crystals were observed one week after setup. Of the 288 precipitant conditions tested, 0.1 *M* sodium citrate buffer pH 5.5, 28% PEG MME 2K and 0.33% LDAO yielded diffraction-quality crystals.

2.4. Data collection

A rhomboid crystal of approximately 50 µm in size in all dimensions was harvested in a Hampton cryoloop and immediately flash-cooled in liquid nitrogen. The cryopin was placed in a puck of the Advanced Light Source (ALS) storage and transfer system (Rupp *et al.*, 2002) and robotically mounted on ALS beamline 5.0.3. Data to 1.65 Å were collected at an X-ray wavelength of 1.000 Å, integrated using *HKL*2000, and scaled with *SCALEPACK* (Otwinowski & Minor, 1997) in trigonal Laue group $\overline{3}$. The data were further reduced in final space group $P3_221$ (No. 154), with unit-cell parameters a = 64.91, c = 87.20 Å. Calculation of Matthews probabilities (Matthews, 1968; Kantardjieff & Rupp, 2003) and solvent density indicated there to be one molecule in the asymmetric unit. Data-collection statistics are summarized in Table 1 and details are available from the PDB header. Successful molecular replacement established $P3_221$ as the correct selection from the enantiomorphic pair (No. 154 *versus* No. 152).

2.5. Structure determination

The structure of MTB RmlC was determined by molecular replacement using a homology model built with the automated protein-structure prediction system AS2TS (Amino-acid Sequence to Tertiary Structure) developed at LLNL (Rupp *et al.*, 2002). The dTDP-4-dehydrothamnose 3,5-epimerase RmlC from the archaeon *Methanobacterium thermoautotrophicum* (PDB code 1ep0, chain *A*; 183 residues; Christendat *et al.*, 2000), which shares an identity of 38% over 185 residues with MTB RmlC (Fig. 1), was used as a template to calculate

Figure 1
Pairwise structural alignment of homologous protein chains with RmlC from MTB using the local global alignment program *LGA* (Zemla, 2003). The left column indicates the PDB code and chain ID and colored bars represent $C^{\alpha}-C^{\alpha}$ distance deviation between superimposed PDB structures and RmlC (200 residues; from the N-terminus at the left to the C-terminus at the right). Residues superimposed below 1.5 Å are in green, below 3.0 Å in yellow, below 4.5 Å in orange, below 6.0 Å in brown and residues at or above 6.0 Å are in red. Non-aligned terminal residues are in gray. The right column contains r.m.s.d.s in Å calculated for all C^{α} pairs that are superimposed under 5 Å distance cutoff. The plot shows that homologous proteins differ significantly (red) from TB RmlC in the C-terminal part (loop 160–165, region 179–186) and also that the C-terminal helix is not present (gray) in the templates.

the main-chain atoms in the model; side-chain atoms were calculated using the program *SCWRL* (Canutescu *et al.*, 2003). *EPMR* (Kissinger *et al.*, 1999) was used with default settings (15–4 Å, no bump restraints), and searches converged at correlation coefficients of 0.33, with an R value of 0.48 after rigid-body refinement against data to 2.8 Å.

2.6. Model building and refinement

To ensure effective phase-bias removal, the model was iteratively built using the program *XFIT* (McRee, 1999) into maps generated by the *Shake&wARP* procedure implemented in the TB consortium map-improvement server (Reddy *et al.*, 2003). An additional C-terminal helix was clearly visible in the initial map, as were several major loop arrangements and residue modifications. After repeated cycles of solvent building and real-space refinement, followed by restrained *REFMAC5* maximum-likelihood refinement (Murshudov *et al.*, 1997), the final structure (PDB code 1upi) refined to R = 0.201 and $R_{\rm free}$ = 0.249. Weak density for three additional residues from the N-terminal His tag was visible in the maps, but these could not be reliably modeled and have been omitted from the model. At the C-terminus, the last two residues from the protein and three additional residues from a *Bam*H1 cloning artefact were also absent. Real-space correlation coefficient plots ($\langle CC \rangle$ = 0.92) have been calculated by the TB consortium map-improvement server (Reddy *et al.*, 2003). Details of the refinement and data-collection statistics are tabulated in the header file of PDB entry 1upi and are briefly summarized in Table 1.

2.7. Quality assessment

PROCHECK (Laskowski *et al.*, 1993) and *WHAT_CHECK* (Hooft *et al.*, 1996) reports were created upon coordinate deposition and are available from the PDB for entry 1upi. Ramachandran plot distribution, coordinate error, r.m.s.d. from target-geometry values and real-space correlation which are typical for a well refined 1.7 Å structure are summarized in Table 1. The only residue in a generously allowed region of the Ramachandran plot region is Ala161, which is located in a disordered loop.

2.8. *In silico* virtual ligand screening

To virtually screen for potential inhibitors of MTB RmlC, flexible docking simulations were performed with *ICM-Pro* 3.0.25l (Schapira *et al.*, 2003). To test the robustness of the docking procedure, crystal structures of RmlC ligand complexes from *Streptococcus suis*, *Salmonella typhimurium* and *Methanobacterium thermoautotrophicum* were simulated first. Similar substrate analogs as well as reported active compounds (Andres *et al.*, 2000; Ma *et al.*, 2001; Babaoglu *et al.*, 2003) were then docked to MTB RmlC using superposition with a ligand-bound structure and inferences from sequence alignment to initially define the receptor site. The active sites in the MTB RmlC dimer were also characterized using ICMPocketFinder, which detects cavities of sufficient size to bind 'druggable' molecules.

3. Results

3.1. Structure summary

MTB RmlC is an obligate homodimer in which the dimer axis in 1upi coincides with the crystallographic twofold. The buried interface excludes 1500 Å^2 per molecule (Fig. 2). The overall topology of MTB RmlC is consistent with the structural classification of these proteins, mainly β-class with a jelly-roll-like topology, in which each monomer is characterized by a double-stranded β-helix forming the active site of the enzyme (Laskowski, 2001). As seen in the other dimer structures, β-strands extending from each monomer stabilize the dimer. The extended strand from the N-terminus of one monomer contributes residues to the active site of the other (Christendat *et al.*, 2000; Giraud *et al.*, 2000; Babaoglu *et al.*, 2003). The most significant difference between MTB RmlC and the other RmlC structures is a well defined C-terminal ten-residue helix extension. The surface-accessible cysteine residues Cys134 and Cys146 could be clearly modeled as *S*,*S*-(2-hydroxyethyl)-thiocysteine (CME; Fig. 3), presumably resulting from modification by β-mercaptoethanol present in the dialysis buffer. The buried cysteine residues Cys50 and Cys76 are not modified.

The structure contains two non-proline *cis*-peptides, Gly60–Leu61, which is clearly defined (Fig. 4), and Asp161–Gly162, which is located in a disordered loop region. *Cis*-peptide Gly60–Leu61, part of the highly conserved sequence VLRGLH, has been observed in all reported RmlC structures and is likely to have biological relevance (Weiss & Hilgenfeld, 1999), as it forms part of the active-site binding pocket. It has been proposed that because this energetically unfavorable conformation is highly conserved, Gly60 may help to orient catalytic residues on β6 in the active site (Christendat *et al.*, 2000). An exception to this *cis* conformation is found in 1dzr (*Salmonella typhimurium*), in which the same residues display a large deviation from geometry targets, indicating uncertainty

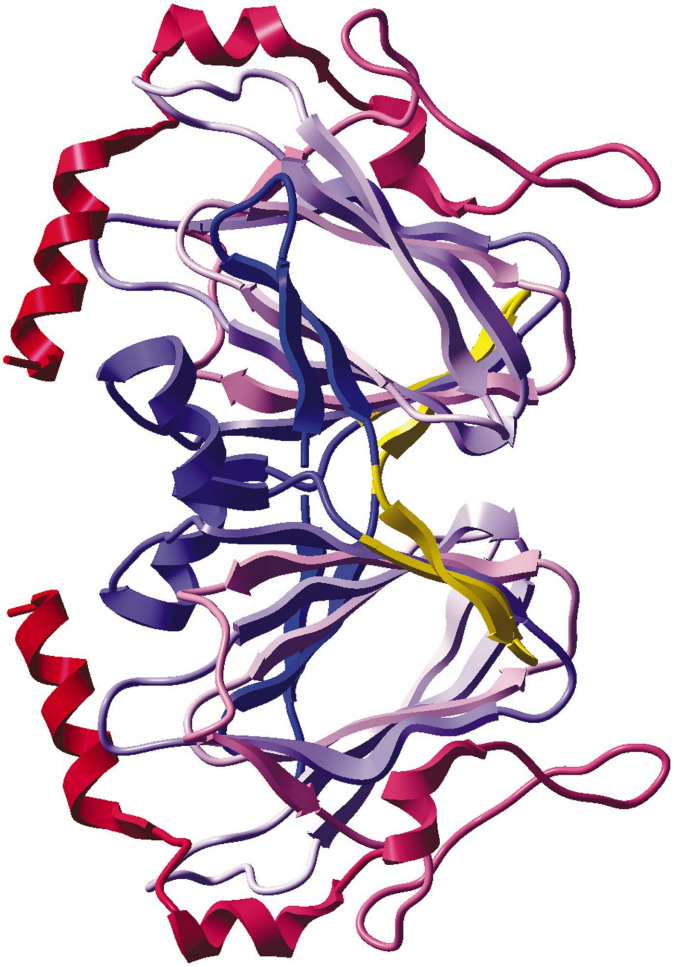

Figure 2
MTB RmlC homodimer. Ribbon drawing showing the homodimer and jelly roll-like topology, in which each monomer (colored by residue number and varying from blue at the N-terminus to red at the C-terminus) is characterized by a double-stranded β-helix forming the active site of the enzyme (Laskowski, 2001). An extended β-strand (shown in yellow) from the N-terminus of each monomer contributes residues to the active site of the other (Christendat *et al.*, 2000; Giraud *et al.*, 2000; Babaoglu *et al.*, 2003). The view looks into the opening of active site of the lower monomer. The image was rendered with *ICM-Pro* v.3.0.25l.

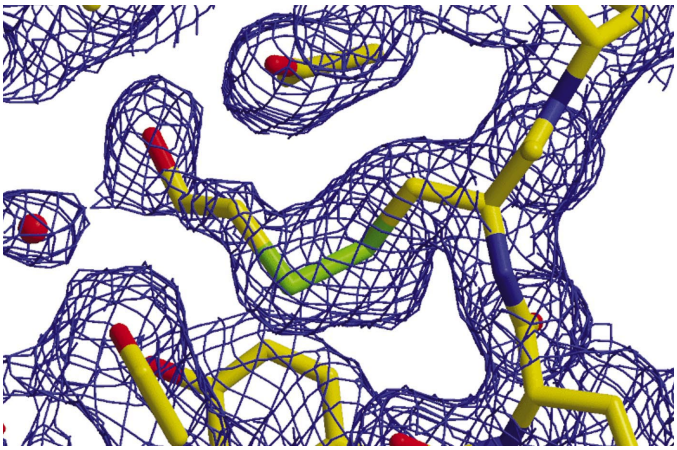

Figure 3
Modified solvent-exposed *S*,*S*-(2-hydroxyethyl)-thiocysteine (CME). The electron-density map around CME134 was generated by the TB consortium bias-removal and map-improvement server (Reddy *et al.*, 2003) and is contoured at the 0.7σ level. The figure was created with *XFIT* (McRee, 1999) and rendered with *RASTER3D* (Merritt & Bacon, 1997).

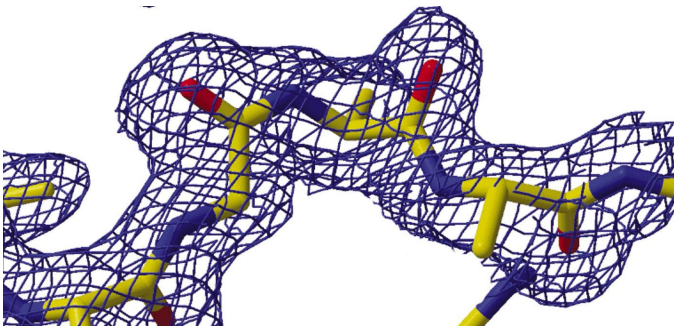

Figure 4
The structurally and sequentially conserved *cis*-peptide Gly60–Leu61 in MTB RmlC (PDB code 1upi). Electron density was calculated by the TB consortium map-improvement server (Reddy *et al.*, 2003) and is contoured at 1σ. The side chains of Leu61 and His62 are clipped for better backbone clarity. The figure was created with *XFIT* (McRee, 1999) and *RASTER3D* (Merritt & Bacon, 1997).

Table 2
Active-site binding-pocket characteristics for MTB RmlC.

PDB code	Volume (Å^3)	Area (Å^2)	Chain *B* residues	Chain *A* residues
1upi chain *A*	541	486	19, 23, 26, 28	47, 49, 51, 59, 62, 70, 72, 119–121, 132, 134, 138, 140, 143–145, 168, 170–171, 174–175
1dzr chain *A*	588	757	2, 15, 17–18, 20–21, 24	27, 29–32, 35–36, 74, 111; 48, 52, 54–55, 58, 60–61, 63, 65–66, 73, 120–122, 131, 133, 139, 144–146, 167, 169–170, 174, 178–179, 181–182
1ep0 chain *A*	535	456	3, 22–24, 26, 28–29, 31	51, 53, 61–62, 64–66, 71, 73, 120–122, 133, 139, 144–145, 169, 171–172, 175
1nxm chain *A*	392	420	29–31, 33–36, 38	61, 63, 65–68, 71, 73–74, 76, 78, 80, 82, 127–129, 138, 140, 142, 146, 175

in the modeling. Inspection of the bias-minimized electron density obtained from the TB consortium bias-removal server (Reddy *et al.*, 2003) suggests that Gly60–Leu61 in 1dzr could in fact be modeled in a *cis* conformation. The *cis*-peptide detection program *ANGL_ANAL* (Weiss & Hilgenfeld, 1999) also ranks the Gly60–Leu61 peptide bond in 1dzr as *cis* with a high probability score.

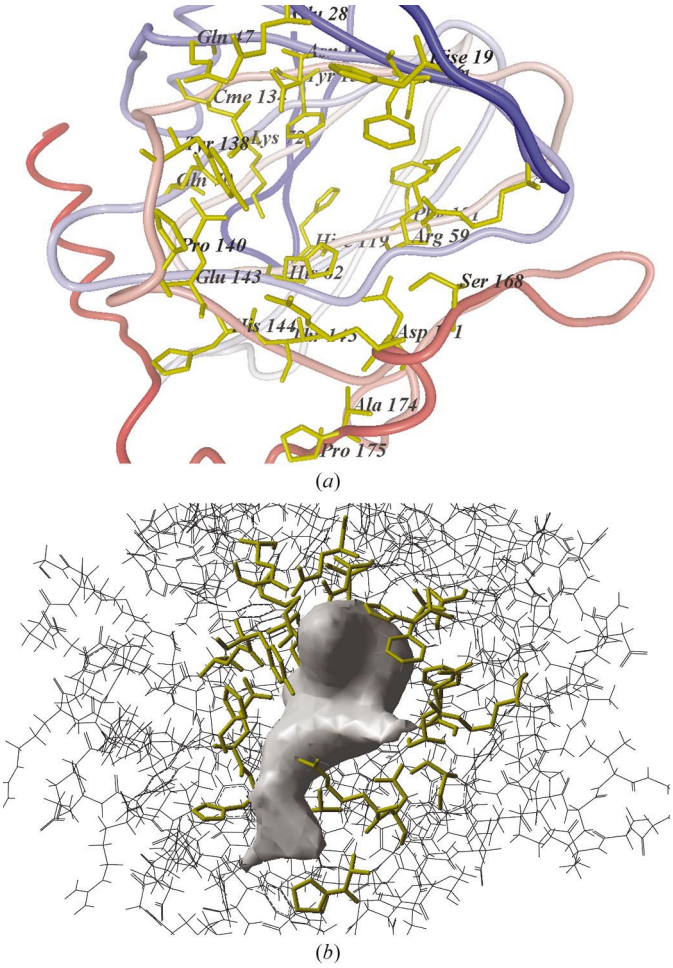

Figure 5
Active site of MTB RmlC. (*a*) Binding-pocket residues as indicated by ICMPocketFinder and reported in Table 2. The polypeptide chain is colored from blue at the N-terminus to red at the C-terminus for each monomer. Yellow sticks represent protein residues describing the pocket. The view is the same as in Fig. 2. (*b*) Binding-pocket volume as defined by ICMPocketFinder for virtual ligand screening. The gray mesh denotes the ligand-accessible pocket volume (500 Å^3) of the protein receptor. Yellow sticks represent protein residues describing the pocket. The view is the same as in (*a*).

Loop Leu159–Ala163 is also disordered in a 2.2 Å MTB RmlC dimer structure (PDB code 1pm7, released during revision of this manuscript) and presents the only major difference between the two TB structures. The total backbone C$^\alpha$-atom r.m.s.d. between each of the two monomers in 1pm7 and the single monomer unit in 1upi excluding Met1, Leu159–Ala163 and the C-terminal Gly197–Met199 are 0.30 and 0.26 Å, respectively. Omitting the same residues in a superposition of the two chains of 1pm7 onto each other yields an r.m.s.d. of 0.16 Å, less than the expected Cruickshank coordinate error (Cruickshank, 1999) of 0.23 Å (1pm7). Together with the very small deviations of the superposition (NCS) matrix from a pure twofold operator perpendicular to axis *c* (0.04°, 0.16 Å displacement along *c*), we conclude that the two structures are highly related. It is also consistent that the higher symmetry structure (1upi) diffracts to higher resolution (1.7 *versus* 2.2 Å in 1pm7). The possibility of significantly different dimer orientations in the two structures can also be excluded, as both dimers superimpose with a total C$^\alpha$ r.m.s.d. of 0.35 Å. Different protein constructs, purification buffers and crystallization conditions are possible sources of the differences in crystal form.

Details of the MTB RmlC active site are shown in Fig. 5. Highly conserved within the MTB RmlC active site are the His–Asp dyads His119–Asp83 and His62–Asp171 as well as Lys72, all of which are ionizable groups strategically placed to participate in acid/base chemistry within the active site (Christendat *et al.*, 2000). Phe121 and Tyr138, which have been implicated in carbohydrate binding (Babaoglu *et al.*, 2003), and the hydrophilic residues Gln47, Asn49, Ser1 and Ser53, which comprise a network for substrate binding and catalysis (Christendat *et al.*, 2000), are also highly conserved. Table 2 reports the characteristics of the active sites in available RmlC dimers, as determined by ICMPocketFinder. The greatest differences are seen in 1dzr (*Salmonella typhimurium*), the total backbone C$^\alpha$-atom r.m.s.d. of which deviates significantly from other RmlCs and the active-site binding pocket of which is distinctly smaller in volume than the other apo RmlC structures. The binding pocket of MTB RmlC is intermediate in volume and surface area between 1dzr (*Salmonella typhimurium*, smallest) and 1nxm (*Streptococcus suis*, largest) and is most similar to 1ep0 (*Methanobacterium thermoauto-*

trophicum). Yet, despite the observed differences in structural details, the sequence and structural elements essential to function are highly conserved among RmlCs, particularly within the double-stranded β-helix forming the active site of the enzyme (Fig. 6).

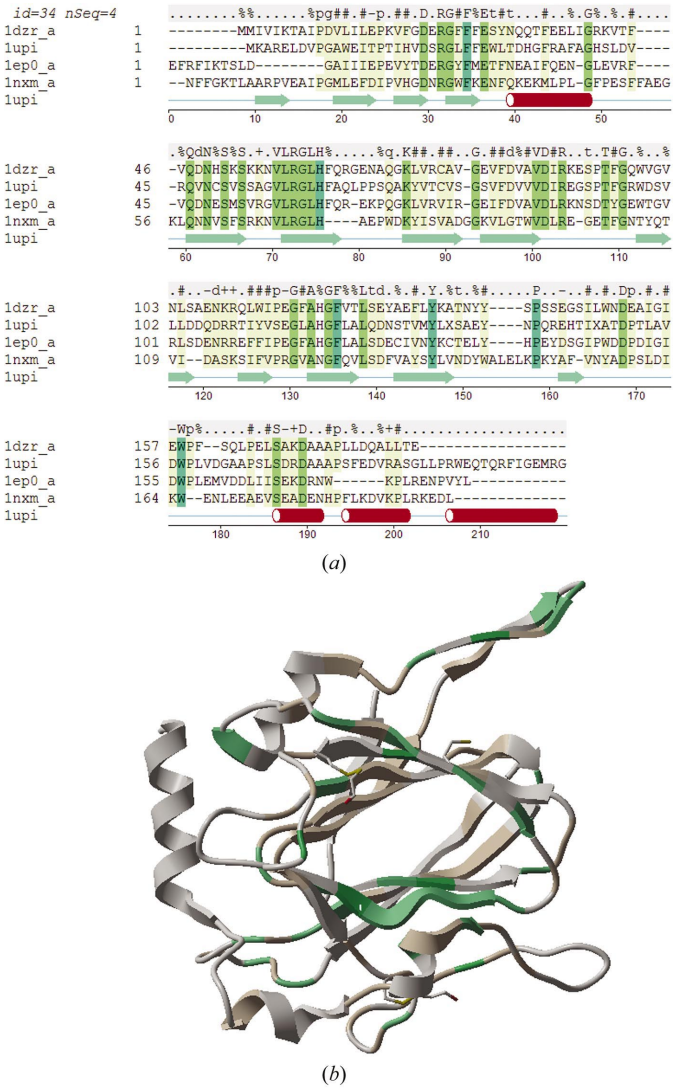

Figure 6
Combination sequence and structural alignment colored by consensus strength (strong in dark green to weak in white). (*a*) Sequence alignment of available RmlC monomer structures. The secondary-structure scheme for MTB RmlC is shown below the alignment and the consensus sequence is shown above the alignment. The consensus string contains the following symbols: +, positively charged amino acids R and K; −, negatively charged amino acids D and E; (^) small amino acids A, S and G; %, aromatic residues F, Y and W; #, hydrophobic amino acids F, I, L, M, P, V and W; ~, polar amino acids C, D, E, G, H, N, Q, S, T, Y; dot, no consensus, no gap. For example, if Gly is found in more than 85% of sequences its consensus symbol is 'G'; if the percentage is between 60–85 the symbol becomes 'g'; if no consensus is established, the symbol becomes '.'. For residues WLVIMAFCYHP, '#' indicates that the residue is found in more than 85% of the sequences and '%' if the percentage is between 60 and 85%. (*b*) Ribbon diagram of MTB RmlC monomer, colored by consensus strength as in (*a*). View is from the bottom of the monomer, the same view as in Fig. 2. Calculated and rendered with *ICM-Pro* (Abagyan *et al.*, 1994, 1997).

3.2. Virtual screening

The ligand conformations in crystal structures of the dimeric RmlC–ligand complexes (*Streptococcus suis*, *Salmonella typhimurium* and *Methanobacterium thermoautotrophicum*) could be reproduced *in silico* within 2–3 Å r.m.s.d., which is considered to be the benchmark for successful simulated docking (Totrov & Abagyan, 2001). Subsequent simulated docking of the same ligands to MTB RmlC produced conformations that closely resembled those observed in the crystal structures of homologs. These simulated complexes were consistently among the top-scoring conformations and were distinguishable from less specific binding modes in that the pyrimidine rings tended to be superimposable; the substrate-binding mode exhibited by these enzymes is a ring π-stacking (Christendat *et al.*, 2000; Giraud *et al.*, 2000; Babaoglu *et al.*, 2003) as seen in the nucleotide-binding regions of many other proteins. Furthermore, although the sugar moiety moves within the active site, as has been observed in crystal structures (Babaoglu *et al.*, 2003), the diphosphates locate themselves similarly, interacting with several conserved arginines. The simulated docking results suggest that the relative superposition of putative inhibitors/analog compounds with pyrimidine and phosphate functionalities should be important factors for drug design. Indeed, ICM successfully docked reported active classes of compounds (Ma *et al.*, 2001; Babaoglu *et al.*, 2003),

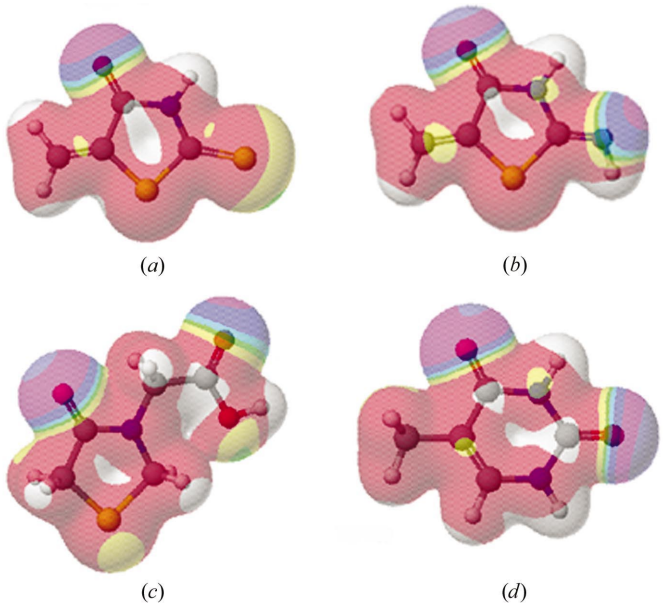

Figure 7
Core structure electron-density isosurfaces for antimycobacterial compounds and substrate analogs colored by electrostatic potential. The potential at a point near a molecule is the potential energy of a positive charge at that point. Coloring is from white (highest, positive) to purple (lowest, negative) calculated with *CaChe WorkGroup Pro* v.6.1.1 at the B88-LYP DFT level (Becke, 1988; Lee *et al.*, 1988). (*a*) Rhodanine core; (*b*) rhodanine-like core; (*c*) thiazolidinone; (*d*) thymine. Compounds with rhodanine or thiazolidinone core structures (*a* and *c*) tended to dock at or near the phosphate-binding region of MTB RmlC, whereas compounds with rhodanine-like core structures (*b*) docked in or near the nucleotide-binding pocket (*d*).

including several compounds from a limited Nanosyn library meeting Lipinski's criteria (Lipinski *et al.*, 2001), with surrogate pharmacophores correspondingly disposed in high-scoring orientations.

As a follow-up to the work of Ma *et al.* (2001), we docked to MTB RmlC a series of compounds containing a rhodanine (van der Helm *et al.*, 1962) or rhodanine-like core structure, which had shown activity against MTB RmlC and inhibited growth of MTB in culture and which the authors had suggested to provide preliminary structure–activity relationships. Consistently, the rhodanine-like core structure bound in or near the nucleotide-binding region by analogy with homologous structures, whereas rhodanine and thiazolidinone core structures bound at or near the phosphate-binding region. The binding preferences for these structures may be explained by their electrostatic potential surfaces (Fig. 7). Substituted thiazolidinones share common features with compounds thought to mimic the diphosphate moiety in the transition state (Biller *et al.*, 1991; Traxler *et al.*, 1991; Prashad, 1993; Barber *et al.*, 1999; Andres *et al.*, 2000; El Zoieby *et al.*, 2003). While this simulated docking has provided additional insights into structure–activity relationships, crystal structures of MTB RmlC complexes will be needed to validate these molecular interactions. Structure-guided *in silico* techniques have been used to rationalize and prioritize compound library design, but not all compounds showing RmlC activity have demonstrated activity against *Mycobacterium tuberculosis* in culture (Ma *et al.*, 2001; Babaoglu *et al.*, 2003). The permeability barrier of the mycobacterial envelope and new routes for drug delivery must also be considered, exploiting functionalities known to exhibit 'good' deliverability in library design and optimization of lead structures.

4. Conclusions

Although the architectures of the active sites of the RmlC enzymes from *Mycobacterium tuberculosis*, *Streptococcus suis*, *Salmonella typhimurium* and *Methanobacterium thermoautotrophicum* are conserved, there are notable differences that emerge based on structural alignment and pocket characterization which contribute to variation in binding-pocket size and shape. In the absence of published crystal structures of MTB RmlC–ligand complexes, such differences in active-site pocket character have a significant impact on structure-guided drug-design approaches that exploit *in silico* protein-ligand docking and underscore the need for ligand-bound crystal structures that more accurately define the protein receptor for simulated docking analysis and elucidate structure–activity relationships.

We acknowledge Uhn Soo Cho, Susan Wachocki, Minyoung So and Min-Young Kim for technical assistance with cloning and protein expression and purification at the LANL TB consortium protein-production facility, Fu Ming Tao at CSU Fullerton for helpful discussions and Clare Smith, Texas A&M University, for critical review of the manuscript. KAK thanks the California State University Program for Education and Research in Biotechnology and the W. M. Keck Foundation for support of the Center for Molecular Structure. LLNL is operated by University of California for the US DOE under contract W-7405-ENG-48. This work was funded by NIH P50 GM62410 (TB Structural Genomics) center grant. AZ was supported by LLNL LDRD grant 02-LW-003.

References

Abagyan, R., Batalov, S., Cardozo, T., Totrov, M., Webber, J. & Zhou, Y. (1997). *Proteins*, **29**, 29–37.
Abagyan, R., Totrov, M. & Kuznetsov, D. (1994). *J. Comput. Chem.* **15**, 488–506.
Andres, C. J., Bronson, J. J., D'Andrea, S. V., Deshpande, M. S., Falk, P. J., Grant-Young, K. A., Harte, W. E., Ho, H. T., Misco, P. F., Robertson, J. G., Stock, D., Sun, Y. & Walsh, A. W. (2000). *Bioorg. Med. Chem. Lett.* **10**, 715–717.
Babaoglu, K., Page, M. A., Jones, V. C., McNeil, M. R., Dong, C., Naismith, J. H. & Lee, R. E. (2003). *Bioorg. Med. Chem. Lett.* **13**, 3227–3230.
Barber, A. M., Hardcastle, I. R., Rowlands, M. G., Nutley, B. P., Marriott, J. H. & Jarman, M. (1999). *Bioorg. Med. Chem. Lett.* **9**, 623–626.
Becke, A. D. (1988). *Phys. Rev. A*, **38**, 3098–3100.
Biller, S. A., Forster, C., Gordon, E. M., Harrity, T., Rich, L. C., Maretta, J. & Ciosek, C. P. (1991). *J. Med. Chem.* **34**, 1912–1914.
Canutescu, A. A., Shelenkov, A. A. & Dunbrack, R. L. Jr (2003). *Protein Sci.* **12**, 2001–2014.
Christendat, D., Saridakis, V., Dharamsi, A., Bochkarev, A., Pai, E., Arrowsmith, C. H. & Edwards, A. E. (2000). *J. Biol. Chem.* **275**, 24608–24612.
Cruickshank, D. W. J. (1999). *Acta Cryst.* D**55**, 583–601.
El Zoieby, A., Sanschagrin, F. & Levesque, R. C. (2003). *Mol. Microbiol.* **47**, 1–12.
Engh, R. A. & Huber, R. (1991). *Acta Cryst.* A**47**, 392–400.
Giraud, M.-F., Leonard, G. A., Field, R. A., Berlind, C. & Naismith, J. H. (2000). *Nature Struct. Biol.* **7**, 398–402.
Helm, D. van der, Lessor, A. D. Jr & Merritt, L. L. Jr (1962). *Acta Cryst.* **15**, 1227–1232.
Hooft, R. W. W., Vriend, G., Sander, C. & Abola, E. E. (1996). *Nature (London)*, **381**, 272–272.
Kantardjieff, K. A. & Rupp, B. (2003). *Protein Sci.* **12**, 1865–1871.
Kantardjieff, K. A. & Rupp, B. (2004). In the press.
Kissinger, C. R., Gelhaar, D. K. & Fogel, D. B. (1999). *Acta Cryst.* D**55**, 484–491.
Laskowski, R. A. (2001). *Nucleic Acids. Res.* **29**, 221–222.
Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
Lee, C., Yang, W. & Parr, R. G. (1988). *Phys. Rev. B*, **37**, 785–789.
Lipinksi, C. A., Lombardo, F., Dominy, B. W. & Feeney, P. J. (2001). *Adv. Drug. Deliv. Rev.* **46**, 3–26.
Ma, Y., Stern, R. J., Scherman, M. S., Vissa,V. D., Yan, W., Jones, V. C., Zhang, F., Franzblau, S. G., Lewis, W. H. & McNeil, M. R. (2001). *Antimicrob. Agents Chemother.* **45**, 1407–1416.
McRee, D. E. (1999). *J. Struct. Biol.* **125**, 156–165.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
Merritt, E. A. & Bacon, D. J. (1997). *Methods Enzymol.* **277**, 505–524.
Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst.* D**53**, 240–255.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **267**, 307–326.
Prashad, M. (1993). *Bioorg. Med. Chem. Lett.* **3**, 2051–2054.
Ramaswamy, S. & Musser, J. M. (1998). *Tuber. Lung Dis.* **79**, 3–29.
Reddy, V., Swanson, S., Sacchettini, J. C., Kantardjieff, K. A., Segelke, B. & Rupp, B. (2003). *Acta Cryst.* D**59**, 2200–2210.

Rupp, B., Segelke, B. W., Krupka, H. I., Lekin, T. P., Schafer, J., Zemla, A., Toppani, D., Snell, G. & Earnest, T. E. (2002). *Acta Cryst.* D**58**, 1514–1518.
Schapira, M., Abagyan, R. & Totrov, M. (2003). *J. Med. Chem.* **46**, 3045–3059.
Schroeder, E. K., de Souza, O. N., Santos, D. S., Blanchard, J. S. & Basso, L. A. (2002). *Curr. Pharm. Des.* **3**, 197–225.
Segelke, B. W. (2001). *J. Cryst. Growth*, **232**, 553–562.
Stern, R., Lee, T., Lee, T., Yan, W., Scherman, M., Vissa, V., Kim, S., Wanner, B. & McNeil, M. (1999). *Microbiology*, **145**, 663–671.
Terwilliger, T. C. *et al.* (2003). *Tuberculosis*, **83**, 223–249.
Totrov, M. & Abagyan, R. (2001). *The Thermodynamics of the Drug–Receptor Interaction*, edited by R. B. Raffa, pp. 603–624. New York: John Wiley & Sons.
Traxler, P. M., Wacker, O., Bach, H. L., Geissler, J. F., Kump, W., Meyer, T., Regenass, U., Roesel, J. L. & Lydon, N. (1991). *J. Med. Chem.* **34**, 2328–2337.
Weiss, M. S. & Hilgenfeld, R. (1999). *Biopolymers*, **50**, 536–544.
Zemla, A. (2003). *Nucleic Acids. Res.* **31**, 3370–3374.